Whole genome sequencing

Torpdahl, Mia; Löfström, Charlotta; Møller Nielsen, Eva

Published in:
Annual Report on Zoonoses in Denmark 2013

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

3. Whole genome sequencing

By Mia Torpdahl (mtd@ssi.dk), Charlotta Löfström and Eva Møller Nielsen

In recent years, new DNA-sequencing techniques have been developed and made available for easy and fairly cheap sequencing of bacterial genomes. These post-'Sanger' sequencing techniques are commonly referred to as 'next-generation-sequencing' (NGS). NGS includes sequencing technologies that have high sequence capacity, but produce short sequences (e.g. as developed by Illumina and IonTorrent), which is compensated for by high sequence coverage (Figure 3.1). This technological development has dramatically reduced the cost of determining the complete, or nearly complete, genetic information for bacterial isolates. The technology and price will in the near future allow reference laboratories to implement whole genome sequencing (WGS) for typing of bacterial isolates as a replacement or complement to the conventional phenotypic and molecular typing methods.

The use of WGS for epidemiological purposes, such as routine surveillance and outbreak investigations, rely on the analysis of the large amount of data produced. Presently, there are two main approaches for analysing data, the mapping of sequence reads to a reference and "de novo" assembly (without a reference). These approaches can be used independently or in combination. The sequences from different bacterial isolates can be compared to each other directly e.g. by identifying single-nucleotide polymorphisms (SNPs) compared to a reference or gene-by-gene comparisons across assembled genomes. Conventional typing information can also be extracted. Currently, the analytical approaches are still under development and no international standards have been decided on for comparing bacterial isolates based on WGS data.

In the last two years, the use of WGS of foodborne bacteria has been under development at Statens Serum Institut and the National Food Institute, Technical University of Denmark, for epidemiological purposes. At both institutes, WGS has been used for several outbreak investigations in 2012-2013 and WGS was tested for the routine surveillance of VTEC infections for a 3-months period in 2012 [1]. Furthermore, Statens Serum Institut implemented WGS for the continuous surveillance of listeriosis in 2013. The analytical approaches used for the routine surveillance might be markedly different from the approach used in outbreak situations where a limited number of closely related isolates are compared, e.g. by SNP-based analysis.

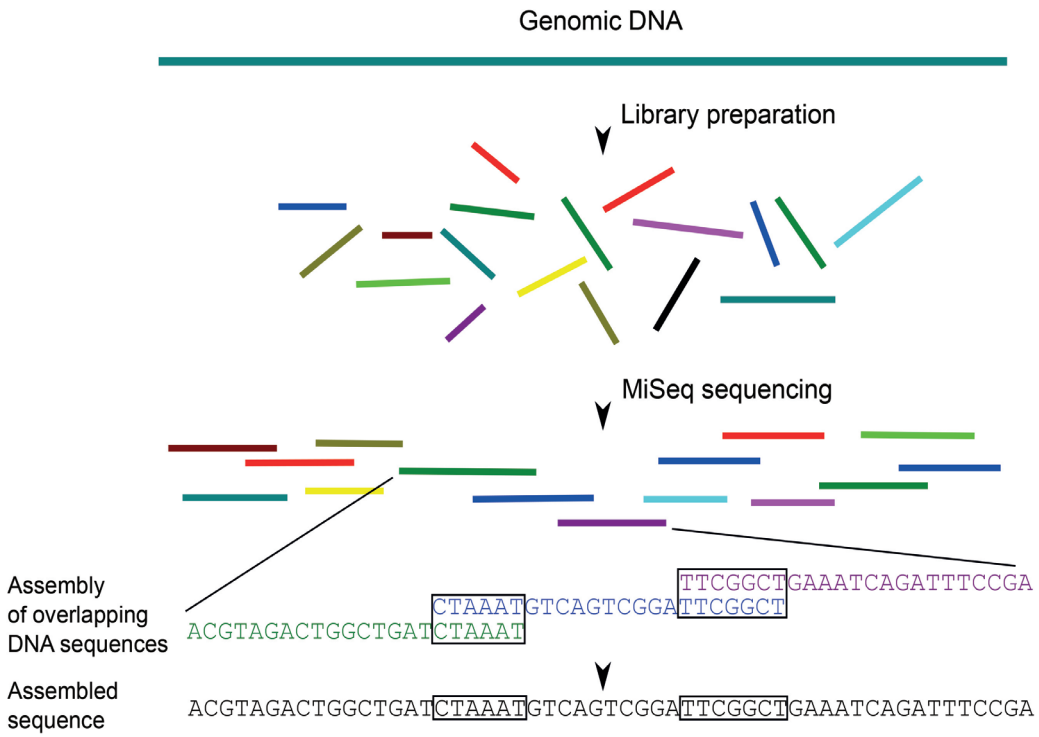
Surveillance of listeriosis

The surveillance of listeriosis cases in Denmark is based on the 'real-time' typing of isolates from all human infections. The regional clinical laboratories submit the isolates to Statens Serum Institut where typing is performed and used for early detection of outbreaks and the surveillance of trends over time. For the past ten years, isolates have been typed by pulsed-field gel electrophoresis (PFGE) (Annual report on Zoonoses 2011). Although PFGE with two restriction enzymes is internationally standardized and gives a good discrimination for typing of *Listeria*, PFGE is relatively labour-intensive to perform on a few isolates at a time and PFGE-profiles can generally be difficult to interpret and compare. As replacement of PFGE and other characterisation methods used for *Listeria* isolates, Statens Serum Institut implemented WGS for surveillance of listeriosis in 2013. The incoming isolates are sequenced on a weekly basis and the aim of the developed pipeline and the analytical approach is to obtain a clear indication of any related isolates and thereby potential outbreaks by an easy and fast analysis. The new isolates should be compared to older isolates in the database without re-analysis of these. Therefore, the analysis is based on extracting the genes from the multiple-locus sequence typing (MLST) method, which has an international standard nomenclature [2]. Only in the situations where isolates with the same MLST type are identified within a few months, a SNP-based analysis is performed to directly compare the relevant isolates. This algorithm for analysis was developed and validated based on the knowledge obtained from analysing representative isolates from the previous ten years' surveillance, including isolates with known epidemiological links.

In 2013, 50 cases of listeriosis were registered. In concordance with the PFGE-based surveillance, WGS showed that the majority of isolates were from sporadic cases. However, five clusters of two to four isolates each were identified. Within these clusters, isolates had the same MLST profile and had less than 10 SNP differences. One cluster seemed to be linked to a specific hospital, but otherwise it was not possible to find a common source of infection or any other link between the patients. Epidemiological investigations of *Listeria* cases are often quite difficult due to the long incubation period, the high mortality rate and the fact that *Listeria* cases are often very old and suffer from underlying diseases.

WGS is a cost-efficient method for typing if used as the only laboratory method, i.e. if the hitherto used methods (serotyping and PFGE) are discontinued. As always when shifting laboratory methods, this can give problems with comparability between laboratories and countries in a transient period. Statens Serum Institut has implemented WGS

Figure 3.1. Principles of next generation short read sequencing



Source: Statens Serum Institut

for typing of *Listeria* isolates from human cases of infection before national or international systems for exchange of WGS-typing data is available. However, the principle of extracting the seven MLST loci with an established nomenclature from the NGS raw data (possible extended with more loci in the future) gives optimal conditions for easy exchange of typing results, as the results are independent of the bioinformatics tools used in each laboratory, e.g. for assembly and SNP-calling.

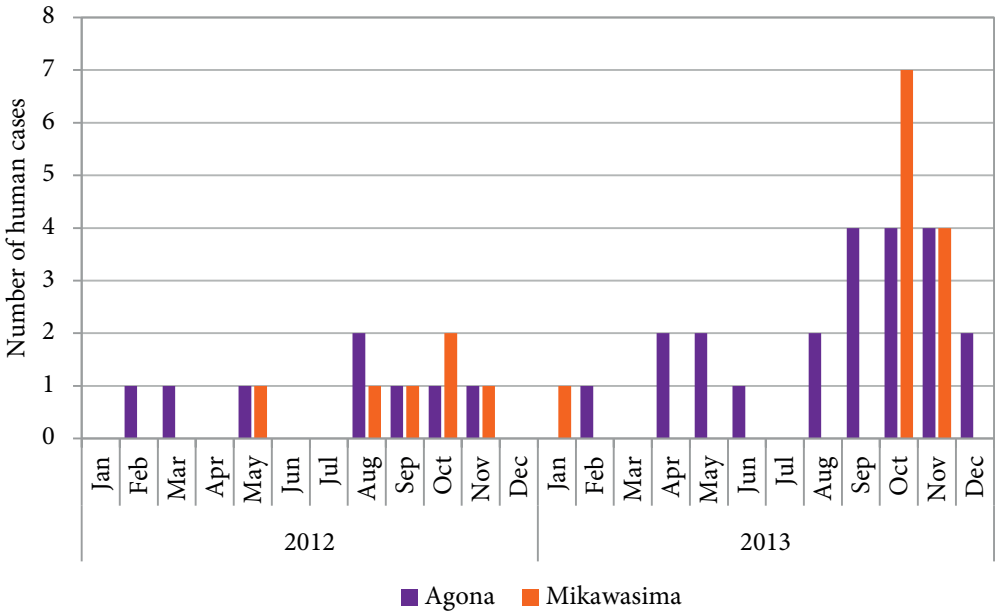
However, PFGE is the standard method for *Listeria* typing in Europe – for typing of human isolates as well as food isolates – a PFGE profile will still be produced for outbreak isolates for national and international comparison, including upload to ECDC's Molecular Surveillance System.

WGS for *Salmonella* cluster investigations

The national surveillance of human *Salmonella* cases is based on the 'real-time' typing of isolates from human infections by serotyping of all isolates, multiple-locus variable number of tandem repeats analysis (MLVA) of the two most frequent serotypes (Enteritidis and Typhimurium, including the monophasic variant) and PFGE of other serotypes. Results are used for cluster analysis, outbreak investigations, comparison with food and animal isolates as well as making international inquiries. There are some interpretative implications that have to be considered when using PFGE in outbreak investigations, e.g. the significance of minor band differences when defining clusters within different *Salmonella* serotypes. In 2013, Statens Serum Institut investigated the use of WGS for cluster detection and analysis of *Salmonella*. The incoming isolates were screened for clusters based on serotype and when potential outbreaks emerged, isolates were selected for sequencing. Two clusters appearing in the autumn were investigated using WGS, one being the rare serotype Mikawasima and the other the more prevalent serotype Agona (Figure 3.2).

Mikawasima isolates from patients in 2012 (6 isolates) and 2013 (12 isolates) were analysed by WGS and PFGE. The 11 isolates from October and November 2013 clustered together by both methods. Analysis of WGS data showed that the isolates in this cluster had no SNP differences, whereas all other isolates clustered into several groups more than 70 SNPs apart (Figure 3.3). The cluster isolates displayed the same unique PFGE profile although the profiles found in Mikawasima isolates generally showed high similarity (Figure 3.3). The data indicate that the increase in Mikawasima seen in October and November 2013 was an outbreak. It also showed another cluster from 2012 consisting of 3 cases. The SNP tree and the PFGE data showed a high degree of concordance, although the sequence data had a higher resolution in comparison to

Figure 3.2. Prevalence of Salmonella serotypes Agona and Mikawasima in 2012 and 2013



Source: Statens Serum Institut

the few band differences seen with PFGE.

The 16 *S.* Agona isolates received from February to November 2013 were analysed by PFGE and WGS. Both PFGE and WGS data identified the same 8 isolates received during August to November as being part of an outbreak. However, WGS data had a higher resolution and separated isolates with indistinguishable PFGE profile in correlation with age group and date of isolations, indicating a unique epidemiological relationship among these isolates.

WGS has also been evaluated as a tool to compare *Salmonella* isolated from humans and foods in outbreak investigations and to investigate the plausibility of sharing data and results between the laboratories at Statens Serum Institut and the National Food Institute, Technical University of Denmark. One example is an outbreak in 2012 where seven patients were infected with *S.* Bareilly with an identical PFGE profile. Patient interviews traced the source to unidentified food served at a specific restaurant [3]. At the same time four broiler flocks were tested positive for *S.* Bareilly. PFGE profiles from the veterinary and human isolates were found to differ by two bands. WGS results showed that the broiler and human isolates belonged to the same MLST type and that they were divided into two closely related groups based on SNPs. It was concluded that there was a close common ancestor of the two isolate groups, but that the broiler flock did not seem to have been the direct source of the human outbreak [4]. Although the same conclusion was drawn based on PFGE and WGS data, a more precise determination of the relationship between isolates was obtained using WGS thereby contributing to the conclusions of the outbreak investigation.

From these first and very promising uses of WGS analysis for cluster investigations on human *Salmonella* infections and outbreak investigations of food and human isolates, we therefore conclude that WGS is a highly sensitive method that will contribute tremendously in cluster detection and outbreak investigations. Over the next few years it is the goal at Statens Serum Institut to replace the current flow of different subtyping schemes in *Salmonella* with one method and therefore implement WGS for *Salmonella* for the national human surveillance.

At the National Food Institute, WGS is planned to be implemented within a few years' time and used as a complement or to replace existing typing methods. In line with this, a PhD project was initiated in 2013 aiming at evaluating different approaches suitable for application of WGS in a routine laboratory setting. In this project cases related to outbreak investigations will first be investigated and thereafter, using the experiences gained, strategies for applying WGS for surveillance of e.g. *Salmonella* will be investigated. During this process it will be important to harmonize the protocols used in order to assure that data can easily be exchanged between SSI and DTU, e.g. in joint outbreak investigations.

WGS for the future surveillance of foodborne infections/zoonoses in DK

The vision for implementing WGS in reference laboratories is that analysis of WGS-data can replace the multitude of different methodologies presently used for characterisation of bacterial isolates, e.g. serotyping, PCR/Sanger sequencing for determining virulence factors and antimicrobial resistance genes, antimicrobial resistance profiling, and molecular typing methods for high discriminatory typing. If the relevant information can be extracted from WGS data, it will be cost-effective to replace the conventional methods by routine WGS in the surveillance of foodborne infections. Statens Serum Institut and the National Food Institute, Technical University of Denmark are working on further developing and implementing WGS for characterisation of foodborne pathogens such as verotoxin-producing *E. coli*, *Salmonella*, *Listeria* and *Campylobacter*. The great benefits of WGS will be achieved when tools are developed for extracting key conventional typing information, primarily the serotype of *Salmonella* and *E. coli*. Conventional serotyping has formed the primary categorization of these organisms for decades and it can be anticipated that serotype information will still be important for a period of several years to come. This will allow the discontinuation of conventional serotyping in laboratories performing WGS as they will still be able to compare to laboratories only performing the conventional serotyping. In contrast, it is less likely that backwards comparability is feasible and important for typing methods like PFGE. To enable sharing of data between different laboratories, it is important to assure that harmonized protocols are applied.

References

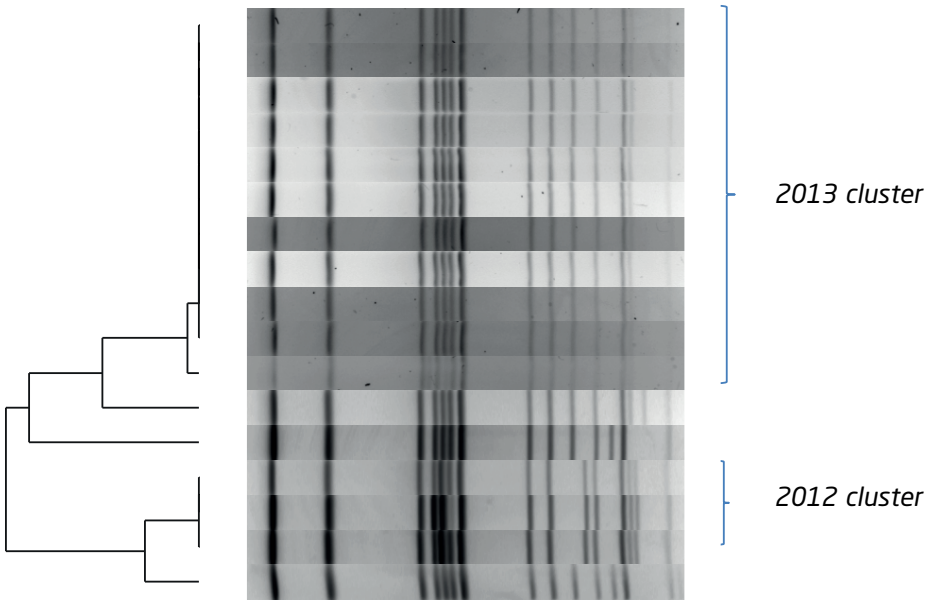
1. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM (2014). Evaluation of Real-Time WGS for Routine Typing, Surveillance and Outbreak Detection of Verotoxigenic *Escherichia coli*. JCM PMID:24574290 (Epub ahead of press).

2. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le Monnier A, Brisse S (2008). A new perspective on *Listeria monocytogenes* evolution. Plos Pathog Plos Pathog doi: 10.1371/journal.ppat.1000146.

3. Anonymous (2013). Annual Report on Zoonoses in Denmark 2012, National Food Institute, Technical University of Denmark.

4. Torpdahl, M, Kiil, K, Litrup, E, Sørensen, G, Aarestrup, FM (2013). A Danish *Salmonella* Bareilly outbreak investigated by the use of whole genome sequencing. In 10th International Meeting on Microbial Epidemiological Markers (IMMEM-10) - Abstract Book. (pp. 215). Paris, France.

Figure 3.3. Dendrogram based on SNPs retrieved from the core genomes of the S. Mikawasima strains. PFGE profiles from the same strains are also shown



Source: Statens Serum Institut